

圖片描述技術輔助文本於可讀性模型準確度之研究

Evaluating the Accuracy of Image Captioning Technology in Enhancing Text Readability Models

Chieh-Hsuan Wu

Graduate Institute of Digital Learning and
Education, National Taiwan University of
Science and Technology
Jessiewu.jjlin@gmail.com

Hou-Chiang Tseng

Graduate Institute of Digital Learning and
Education, National Taiwan University of
Science and Technology
tsenghc@mail.ntust.edu.tw

Kuan-Yu Chen

Department of Computer
Science and Information
Engineering National Taiwan
University of Science and
Technology
kychen@mail.ntust.edu.tw

Berlin Chen

Department of Computer
Science and Information
Engineering, National Taiwan
Normal University
berlin@ntnu.edu.tw

Yao-Ting Sung

Department of Educational
Psychology and Counseling,
National Taiwan Normal
University
sungtc@ntnu.edu.tw

摘要

「適性閱讀」是將文本難度與讀者能力做適配，使讀者可以更容易理解文本內容。然而，隨著書籍的大量出版，由人工進行難易度評判將費時耗力，因而致力發展可讀性模型。過往文本可讀性模型的發展常以文字作為特徵，並未將圖片的內容作為評估文本難度的考量。然而，有其他領域指出，圖片能影響閱讀理解。受益此研究的啟發，本研究將利用圖片描述技術抽取圖片的內容來做為特徵加入文本可讀性模型，以探討圖片內容的加入是否能夠提高可讀性模型的準確度。

Abstract

“Adaptive reading” is the process of adapting the difficulty level of a text to the

ability of the reader so that the reader can understand the text more easily. However, with the large number of books being published, it is time-consuming and labor-intensive to manually assess the difficulty level, and therefore, efforts have been made to develop readability models. In the past, the development of text readability models was often characterized by text, and the content of pictures was not considered as a factor in assessing the difficulty of text. However, it has been suggested in other fields that pictures can affect reading comprehension. Inspired by this research, this study will utilize picture description techniques to extract the content of pictures as features to be added to the text readability model, in order to investigate whether the addition of picture

content can improve the accuracy of the readability model.

關鍵詞：文本可讀性、圖片描述、英文可讀性模型

Keywords: Text Readability, Image Captioning, English Readability Model

1 介紹

閱讀是學習重要的方式之一。閱讀不只能夠開闊視野，超越時空跟古人交談，獲取前人的智慧（曾志朗，2000）；亦能培養不同的能力，如閱讀過程中需要文字辨認與文意理解的能力，對於字詞與文句培養解碼能力；文意解讀方面則培養文本材料處理的能力（Carretti et al., 2020）。閱讀的好處不只是能力的培養，洪蘭（2005）於認知神經科學領域中，閱讀有助於刺激大腦神經的發展，預防大腦的退化；同時，透過閱讀，人們能夠減少因無知而帶來的恐懼，減輕挫折感（洪蘭，2005）；更是減輕壓力（Levine et al., 2022）。從文意理解方面至認知神經科學等，各界皆有學者從事相關研究，探討與閱讀有關之影響力或能力；可見閱讀對於個人的影響不可小覷。

然而，閱讀的重要性不僅對個人產生影響，更會影響到一個國家在國際競爭中的地位（OECD, 2009）。經濟合作暨發展組織（Organisation for Economic Cooperation and Development, OECD）於2009年的報告中指出：人力資本是當今最重要的資本形式，而國家的教育水準則是其經濟成長潛力的預測指標。透過閱讀行為的培養，不僅提升了個人的能力，也間接提高了國家的競爭力（Schleicher et al., 2009）。因此，各國開始重視閱讀能力的培育；並且透過參與國際間舉行的評量來了解各國學生閱讀表現與競爭力（張毓仁等，2011）。

國際學生能力評量計畫（Programme for

International Student Assessment, PISA）針對參與國中的十五歲學生進行閱讀、數學和科學的評量；希望能為完成基礎教育學生的關鍵素養，提供具體的參考資訊（洪碧霞等，2018）。除此之外，另一項針對閱讀素養進行評量的促進國際閱讀素養研究（Progress in International Reading Literacy Study, PIRLS）為國際教育評估協會（The International Association for the Evaluation of Educational Achievement, IEA）所舉辦，為評估四年級學生的閱讀理解能力、閱讀素養成就及寫讀能力；並透過結果研究學生的能力及學習環境的影響（柯華葳等，2016）。由上述可知，兩個評量皆涵蓋閱讀素養的測驗及評估，可見閱讀於現今社會中的重要性；不僅如此，兩者皆藉由國際間評比的方式，提供各國關於教育成效的資訊，期盼能協助各國在制定與修改教育政策時有所依據（洪碧霞等，2018；柯華葳等，2016）。

我國對於閱讀素養的培育同樣重視。教育部於2000年開始推動閱讀教育，撰寫計畫並投入資源；從閱讀資源的普及，從都市至偏鄉的推廣，投入大量圖書資源與人力並且設立圖書館等適合閱讀的環境；至「提升國民中小學學生閱讀教育實施計畫」的三年期計畫（教育部，2015）。此計畫因應「十二年國民基本教育課程綱要」之理念為提升國中小學生閱讀素養（教育部，2018），除了將閱讀資源整合共享、促進學生們參與多元的閱讀行動外；更是推動「自主閱讀適性學習」，當學生透過數位網路取得多元文本或大量訊息的同時，培養資訊素養與辨別能力（教育部，2015）。

為回應適性閱讀的議題，各級學校紛紛推行相關競賽及活動，邀請學生與家長一同加入行列。然而，文本的挑選適合於否，成為重要的議題。因為，面對文本與書籍等大量出版，靠著專家以人工的方式評判文本難易度不僅需要花費大量時間，更是費力且須以龐大開銷支持；

為解決此現象，文本可讀性模型的開發，使得人們可藉著自動化工具大量檢測文本難易度，來處理如此巨量的資料。事實上，文本可讀性的相關研究早在 20 世紀初，便有學者著手相關研究。如 Thorndike (1921) 將整理英文詞彙之使用頻率，並將其作為頻率表，欲提供做為文本難度的標準；Lively 與 Pressey (1923) 則依據該頻率表發展出第一個可讀性公式；Flesch Reading Ease (Flesch, 1948) 將句長以及詞彙音節數納入做為特徵，等等選擇淺層的語言特徵發展可讀性公式的傳統可讀性模型(翁詩諺 等, 2019)；除此之外，亦有使用各種語言特徵與機器學習模型相結合，這些文本被轉換成許多語言特徵，這些特徵被輸入到 Support Vector Machine(SVM)和 Multilayer Perceptron(MLP) 等簡單模型中 (Schwarm & Ostendorf, 2005; Vajjala & Meurers, 2012)；直到現今仍有許多學者持續研究相關領域，可見可讀性模型議題十分重要 (Patel et al., 2023; Murgia et al., 2023)。

由上述可知，目前可讀性模型大多著眼於文字、字義的研究 (Collins-Thompson, 2014)，對於探討圖片資訊對於可讀性的影響尚未有相關研究；然而，在其他領域上已有許多研究證實視覺與閱讀理解之間互有影響力。例如：科學圖文閱讀領域的研究指出，於科學文本中添加圖像會影響讀者的閱讀行為，但無法保證學習表現良好；透過教學介入與適當引導，有助於閱讀理解之提升 (王孜甯、簡郁芬, 2022)；當文字將與之相呼應的圖片作為提示，或許可以減少認知負荷並促進資訊整合。當語言與圖像特徵連結於同一材料時，根據多媒體認知理論，自文字與圖片中學習成果將比只從文字中學習更佳 (Abdulrahman et al., 2020)。由上述研究可知文本中的圖片會影響讀者的理解與閱讀，透過圖片輔助作為提示幫助讀者在文本上的解讀，或許能提高讀者對於文本的理解。本研究受益於上述研究的啟發，在可讀性模型中加入

圖片內容作為特徵，以探討是否能提升可讀性模型的準確率；換句話說，本研究僅探討加入圖片內容對於可讀性模型效能的影響，並非探討將圖片內容的資訊融入原始文章中對於讀者的影響。本研究將使用圖片描述 (image captioning) 技術，從圖片生成語意，作為可讀性模型特徵，並透過回歸模型探討圖片敘述是否能提高模型準確度；期盼證實圖片內容做為特徵能夠提高可讀性準確率。本論文接下來的安排如下：第二章將回顧與文本可讀性模型以及圖片擷取技術相關之文獻；第三章呈現實驗材料與流程相關內容與操作方式；第四章則依實驗之結果進行分析與討論；最後第五章將提出未來可研究之方向。

2 文獻回顧

可讀性 (readability) 是指讀者能夠理解文本的程度 (宋曜廷等, 2013)。至於，可讀性的高低則受文字特徵以及讀者本身差異而有所影響 (Collins-Thompson, 2014)。由於讀者的能力和背景各不相同，對於文本的理解也會有差異。因此，不同領域常使用可讀性模型來評估文本的難易程度，如 Flesch-Kincaid Grade Level (FKGL) 大多用於教育、醫學領域，Lexile 或 Advantage-TASA Open Standard (ATOS) 等則用於商業領域 (CLEAR, 2023)；除以上模型以字詞出現頻率作為可讀性模型特徵外；Mesmer (2020) 表示，Degrees of Reading Power、Reading Maturity Measure (RMM) 和 Source Rater 以及上述提及之 Lexile 與 ATOS，為五種常用的可讀性公式；同樣採用「句長」、「音節數量」、「字詞出現頻率」等與文字、語意相關作為特徵。

可讀性特徵的研發十分多元，自字數等文字表面特徵至語意、上下文句等較深入意義的特徵；密西根大學教授 Collins-Thompson (2014) 將可讀性特徵進行回顧與整理，提出特徵類別

圖，將特徵以低到高分為六個層次：文本易讀性、詞彙/語意、文法、文本結構、高級語意學、讀者興趣與背景；可見大多以文字本身及語意作為特徵；然而，除了文字上的特徵外；亦有其他的研究從排版(layout)的角度出發，如「字體」、「格式」、「間距」等 (Collins-Thompson, 2014)，探討視覺上對於可讀性的影響。

由上可知，過去可讀性模型大多使用文字作為特徵，雖然有部分研究使用視覺相關的特徵，而這些研究也只將圖片排版納入考量 (Collins-Thompson, 2014)。然而，其他領域研究指出圖片內容能輔助讀者的閱讀理解，如王孜甯與簡郁芬 (2022) 證實將圖片加入科學文本中，透過合適的引導幫助讀者增加對圖文理解程度 (王孜甯，簡郁芬，2014)；張菟真與辜玉旻 (2011) 則透過放聲思考法及附圖的自然文本，探討國小學生閱讀策略及插圖運用狀況，發現高閱讀理解能力者能整合圖文訊息且運用插圖提高理解程度，由此可知，圖片內容能幫助閱讀理解。受上述研究的啟發，本研究欲使用圖片描述技術抽取圖片內容，並使用圖片內容做為可讀性模型特徵，了解圖片是否能作為文本輔助，提高可讀性模型準確率。

欲加入圖片內容作為特徵，需事先將圖片內容轉譯成文字。資訊領域中，圖片描述技術可從圖片中擷取特徵，進行圖片辨認，過去做的研究大多是關於符號與文字辨識以及圖片分類 (Wang et al., 2023; Wang et al., 2023)。然而，近年來多模態大型語言模型 (Multimodal Large Language Model, MLLM) 的發展，能夠從輸入多種媒體型態，如圖片、音訊、影片，輸出相對應的圖片或文字 (Huang et al., 2024)；本研究將利用圖片描述技術，擷取圖片內容作為可讀性特徵，期能夠提升可讀性模型效能。

3 實驗

本研究欲使用圖片描述技術，從圖片中擷取

圖片內容成文字，將之視為特徵加入至可讀性模型中，期望可以透過圖片內容的加入提升可讀性模型之準確率。

3.1 實驗資料

本實驗使用由 Crossley 等人 (2023) 提出 CommonLit Ease of Readability (CLEAR, 2023) 資料庫作為實驗材料；此資料庫請專業教師們為 4724 筆資料挑選出合適資料文本之摘要，並透過 Bradley-Terry 模型取得難度；內容依據符合英文母語者三年級至十二年級課程使用為準則做資料收集，摘錄 250 年來兩種不同寫作流派，供更多領域的學者作使用 (Crossley et al., 2023)。並將文本主要分為資訊文本以及文學文本，且將資訊文本細分為：歷史/社會科學、科技以及技術學科等。隨著文本種類的豐富，此資料庫亦收入不同難易程度的文本；本實驗以 Lexile reading band (Smith, 1989) 做為資料篩選之標準，Lexile 難度與學習者年齡對照表如表 1。

年級	學期末	
	50%	90%
幼稚園	BR160L	150L
1 年級	165L	570L
2 年級	425L	795L
3 年級	645L	985L
4 年級	850L	1160L
5 年級	950L	1260L
6 年級	1030L	1340L
7 年級	1095L	1410L
8 年級	1155L	1470L
9 年級	1205L	1520L
10 年級	1250L	1570L
11 年級	1295L	1610L
12 年級	1295L	1610L

表 1. Lexile 難度與學習者年齡對照表

以 Lexile 網頁預設的建議，其中記載以英文為母語學習者於學期初及學期末各百分比學生

之閱讀能力；隨著年紀增加，學生能接受的文本難度也逐漸提高，這與學生的識字程度密切相關。因此，在識字量較少的低年級，文本通常會配有圖片，以幫助學生更好地理解內容。

在此，本研究選取閱讀資料中圖片較多之年級：英文母語學習者一年級至四年級，選取 Lexile 300~900 的資料；於其中挑選閱讀材料含有圖片且圖片內容符合文本文意之資料，共 155 筆資料，作為本次實驗之測試資料；而剩餘未符合上述條件之 4569 筆資料作為訓練資料。

3.2 實驗工具

3.2.1 可讀性特徵：Global Vectors

本研究使用史丹佛大學提出的 Global Vectors 語言模型做為訓練可讀性模型的特徵 (Pennington et al., 2014)。該模型將單字表示為連續空間中的向量，並以 300 維度作為表示；基於全域對數雙線性回歸 (global log-bilinear regression) 與統計特性，計算成本與其他模型相比較為低且易執行；因此，有許多研究使用此語言模型或進行優化 (Kulkarni et al., 2020; Sakkettou & Ampazis, 2020)。本研究使用此語言模型抽取可讀性特徵，將文本中的詞彙，以查表方式找出詞彙的語意向量，並將所有詞彙的向量以累加的方式得出一個 300 維度的文章語意向量來做為特徵。詳細解釋於 3.3 實驗流程當中加以解釋。

3.2.2 視覺-語言模型：ChatGPT-4

本研究使用 OpenAI 開發的 ChatGPT 作為圖片描述生成工具，此模型能根據使用者輸入的內容，利用人工智慧技術自動生成敘述，以回答使用者提出的各種問題，介面採用對話的方式，模擬與真人對話的情況，使情境更為擬真 (Open AI, 2023; Kalla et al., 2023)。過去，ChatGPT 只接受文字輸入，相關研究大多以使用者體驗與 ChatGPT 所帶來的影響作為探討 (Fuchs, 2023; Kalla et al., 2023) 然而，GPT-4

模型的誕生，打破了只能輸入文字的限制。GPT-4 是一個多模態的大型語言模型，基於 Transformer 模型進行訓練，能夠處理圖片及文字的輸入，並生成文字敘述；因此，為研究圖片內容對於讀者理解的影響，本實驗使用了 ChatGPT 中的 GPT-4 模型 (OpenAI, 2023) 來為圖片生成描述。

3.2.3 可讀性模型：Decision Tree Regressor

為更精準地預測文本的可讀性結果，本研究選用 Scikit-learn 函式庫 (Pedregosa et al., 2011) 中的 Decision Tree Regressor 決策樹回歸函數進行模型建置。該函數具有減少記憶體占用、提升計算效率和速度的特性，並且可以通過調整權重來進行不同的實驗，適用於許多不同產業的預測 (Rahul et al., 2013; Joshi et al., 2020)。

3.3 實驗流程

本實驗使用 CommonLit Ease of Readability (CLEAR, 2023) 資料庫作為實驗材料，以 Lexile (Smith, 1989) 做為文本難易度指標選取 Lexile 300~900 之資料，從中選取閱讀材料中涵蓋圖片且圖片內容與文本同源之材料，作為本實驗的測試資料；其餘則作為訓練資料。本實驗將基於圖片影響閱讀理解，透過圖片特徵的有無作為變因，將測試資料分為含有文本及圖片敘述的實驗組以及單純文本的對照組。對照組實驗流程如下圖 1 所示。

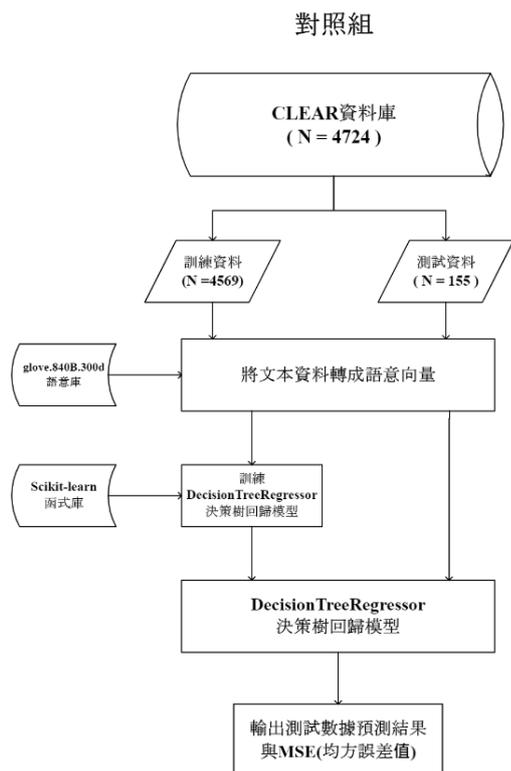


圖 1. 對照組流程圖

將文本測試資料以及訓練資料分別透過程式進行文字轉語意向量的轉換；引用 Global Vectors for Word Representation 概念模型 (Pennington et al., 2014) 中的 glove.840B.300 語言模型，運作過程如圖 2 所示。

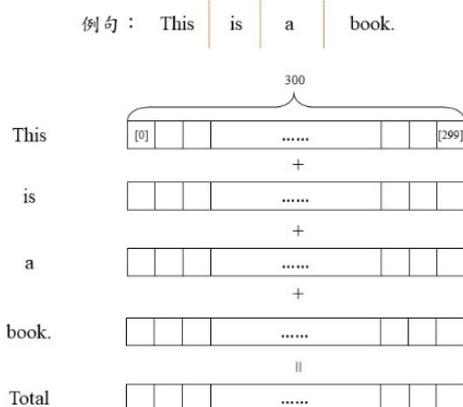


圖 2. 文本轉譯語意向量步驟

每筆資料依空格切分為單詞，逐一讀取每個字詞，至語言模型中搜尋該字詞 300 維度的語意向量並回傳，加到計算文本 300 維度語意向量的矩陣容器中；累加該文本中每一單詞的

300 維度語意向量，直到讀完整篇文章，即可獲得該筆資料總共的語意向量，作為訓練可讀性模型的特徵。

將語意向量轉換完成的資料，加入 CommonLit Ease of Readability (CLEAR, 2023) 中提供的 Bradley-Terry 值作為正確答案，並將訓練資料輸入以 Decision Tree Regressor (Pedregosa et al., 2011) 建立的決策樹回歸模型，進行可讀性模型的建置；最後，將同樣完成語意向量轉換的對照組資料輸入模型進行可讀性預測；即可得到各文本的預測值以及整體與正解的均方根誤差 (Mean Squared Error, MSE)。

原文本加入圖片內容轉述的實驗組流程，與對照組大致相同，實驗流程如圖 3 所示。

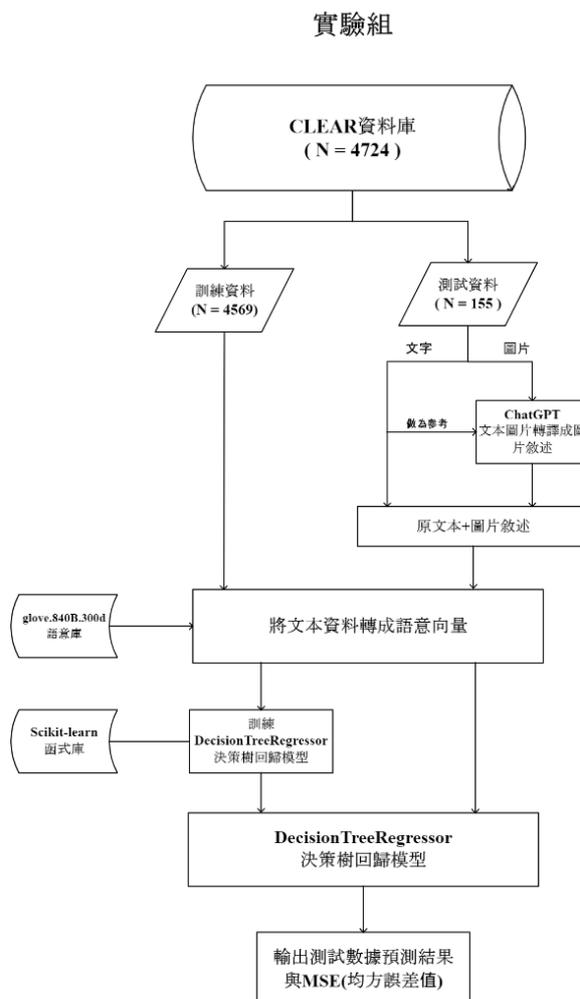


圖 3. 實驗組流程圖

本研究根據圖片敘述的有無作為實驗與對照組的變因，因此實驗組使用 ChatGPT4.0 (Open AI, 2023) 的視覺—語言模型，將原文做為參考，同時輸入圖片取得圖片描述，並將輸出的描述文字合併至原文本最後的位置，做為特徵的增加；其他流程則與對照組相同：使用 glove.840B.300 語言模型將文本轉換為語意向量；並使用與對照組相同回歸模型，輸入測試資料進行預測。將對照組與實驗組結果相比，判斷是否實驗組有低於對照組，驗證當圖片文意加入時是否可以更靠近真實文的文本難度。

4 實驗結果與未來展望

4.1 實驗結果

為證實圖片的內容特徵能提升可讀性模型的效能，本研究將圖片內容透過圖片敘述生成技術，使用視覺-語言模型將圖片內容依照文本生成補述，作為可讀性模型特徵，以提高可讀性模型準確率，實驗結果由表 2 可知。

	MSE 值
實驗組 (文本加入圖片描述)	1.60
對照組 (純文本)	2.59

表 2. 實驗結果 (N=155)

實驗組所預測出與正解之均方誤差為 1.6，對照組則為 2.59，實驗組之均方誤差低於對照組，兩組數值相差 0.99，表明本研究將視覺-語言模型所生成的圖片敘述加入文本的做法，確實能夠提升可讀性模型在文本難度評估上的表現，圖片內容對於提高文本可讀性的貢獻顯著。除此之外，CommonLit Ease of Readability (CLEAR, 2023) 語料庫包含各式學科，由於各學科類型不同，圖片的表徵結構與內容及使用方式也有所差異，因此文本與圖片內容的圖文相關程度也是本篇的探討方向。本研究按語料庫內提供各資料之學科，分類成：History (歷史)、Science (科學)、Technology (科技)、Bio (生物) 以及空格；其中，本研究所使用的未

註明學科資料大多來自於 gutenberg.org 網站，內容以黑白圖文小說居多，其次則來自其他兒童圖書網頁，內容則以文學與科學故事為多。各學科於實驗組與對照組表現如表 3 所示。

學科	篇	實驗組 MSE	對照組 MSE
未註明學科	68	1.57	2.88
History	22	1.47	3.91
Science	47	1.56	2.3
Technology	21	1.95	0.88
Bio	1	1.15	2.25

表 3. 各項學科與各組 MSE 比較 (N=155)

由表知，歷史、科技、生物及未註明學科，MSE 值皆呈現實驗組低於對照組，意謂著在上述學科中，加入圖片內容作為特徵，確實能夠幫助可讀性模型提高準確度；然而，反觀文本類型為科學者，其實驗組之 MSE 值則低於對照組，表現不太理想。

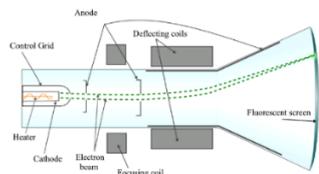
4.2 討論

圖片內容敘述作為特徵加入文本末端，是否有助於提高可讀性模型準確度，可以從整體結果的 MSE 值比較得到證實，擷取圖片內容成為特徵，確實降低模型誤差使預測更精準。表 3 將結果分成各科，由表 3 可知，不同學科會有不同的影響；多數科目除科技外，以 MSE 角度觀察，實驗組皆低於對照組，其中以歷史及未註明學科兩項，實驗組與對照組 MSE 值落差最大、表現最佳。歷史學科多以歷史名畫、偉人雕像居多，其文本資料大多來自維基百科；其內容於網路上有大量的資料，圖片描述過程中，有較多的內容可供參考與生成；未註明學科大多以黑白手繪圖畫居多，來自 gutenberg.org 網站；資料內容大多為 Lexile500，適合英文母語者小學一至二年級程度，內容較為簡單且圖片複雜度低，在圖片描述上有較佳的轉譯結果。

反觀本次表現最差的學科—科技，其資料雖然大多來自維基百科且包含真實照片，但因其

內容包含物品實體圖、相關原理流程圖及示意圖；當讀者在進行閱讀時，圖片的作用大多是使用圖式輔助讀者理解運作過程及使用方式；以資料 Cathode Ray Tube (陰極射線管) 為例，表 4 為該筆資料的圖片及圖片描述。

序號：	Lexile:	主題：
1937	900	Cathode ray tube
圖片		



圖片描述

This diagram illustrates the internal structure of a CRT, showing the electron gun and path to the phosphor screen where the image is displayed.

表 4. 科技學科其一資料圖片以及圖片敘述

此資料文本內容詳細描述陰極射線管的背景及運作方式，然而由表四可知圖片描述唯有指出此圖為陰極射線管，並且以較精簡的文句敘述過程。由此可知，兩者內容雖然相同；但在用詞上，圖片描述卻使用原文並沒有使用的字眼，如：「路徑」(path)、「內部結構」(internal structure) 等，因而產生差異；另外，圖片描述內容十分簡潔，無法真實模擬讀者在閱讀時，對於整個運作流程理解的過程，使圖片描述無法成為文本輔助。

4.3 結論與未來研究

在與閱讀相關的研究中指出，在閱讀資料中圖片能夠影響讀者理解，輔助讀者理解文本內容。然而，過去的可讀性模型所使用的特徵卻大多停留在文字的層次，圖片的資訊鮮少使用。因此，本研究使用圖片描述技術，將圖片內容轉譯並加入原始文本中，進行文本難易度預測。實驗結果顯示額外加入圖片內容敘述的實驗組，MSE 誤差低於純文本的對照組；證實圖片內容

能輔助讀者閱讀，提升讀者對於文本的理解。另一方面，使用圖片內容作為資料特徵，並且加入原有的特徵當中，不只提高可讀性模型準確率，使可讀性模型能更接近專家評判，增加可信度；也因本研究單純將圖片特徵加入至測試資料當中，可讀性模型準確率立即提高，得知不須更改或重新訓練可讀性模型內容，亦可提到準確率；由此可知，本研究所研發之技術有潛力直接應用於其他可讀性模型，也許同樣能夠提高準確率。換言之，無須使用本研究所建立的模型，只需於欲使用的模型中，將測試資料加入新特徵，或許亦可完成任務，使文本難易度的測量更加省時省力。

然而，圖片內容輔助文本理解在各科上影響力略有不同；此現象造成的原因或許牽涉本研究資料量不足，僅能以少量數據做為討論；各學科資料不均，導致無法更深入了解各學科之間圖片輔助文本的影響力，以及文本型態和圖片風格差異所帶來解讀上的影響。

有鑑於此，為更加客觀的了解圖片輔助文本在各學科上，幫助讀者理解的程度，增加資料筆數，抑或提升資料豐富度如：增加學科種類、分析各學科圖片資訊等，是接下來進一部探討的方向；除此之外，不同語料庫及語言模型的選擇，查驗語料庫或語言模型的不同是否影響圖片輔助文本於模型的預測結果亦是未來可研究的方向。期盼能更進一步了解圖片內容做為文本輔助幫助讀者理解與作為新特徵提高可讀性模型準確度相關的內容。

致謝

This work was supported by the National Science and Technology Council of Taiwan under Grants NSTC 113-2410-H-011 -001, NSTC 112-2628-E-011-008-MY3 and NSTC 113-2640-B-002-005. This project was financially supported by the “Empower Vocational Education Research Center”

of the National Taiwan University of Science and Technology (NTUST) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. We thank the National Center for High-performance Computing of the National Applied Research Laboratories (NARLabs) in Taiwan for providing computational and storage resources.

參考資料

王孜甯、簡郁芬 (2022)。科學圖文閱讀眼動研究之系統回顧。國立臺灣師範大學教育心理與輔導學系教育心理學報, 53 (4), 773-800.
[https://doi.org/10.6251/BEP.202206_53\(4\).0001](https://doi.org/10.6251/BEP.202206_53(4).0001)

日本文省部 (2007)。日本文省部科學省。
<https://www.mext.go.jp/>

柯華葳、張郁雯、劉子鍵、詹益綾、陳家興、賴明欣 (2016)。台灣國際閱讀素養調查--PIRLS 2016 (第4年)

曾志朗 (2000)。閱讀是多元智慧成功的基本條件。教師天地, 106, 4-5。

洪蘭、曾志朗 (2000) 兒童閱讀的理念— 認知神經心理學的觀點。現代教育論壇— 兒童閱讀的理念與策略, 國立教育資料館, 台北市立師範學院, 2000-11-15, 2000.

張毓仁、柯華葳、邱皓政、歐宗霖、溫福星。 (2011)。教師閱讀教學行為與學生閱讀態度和閱讀能力自我評價對於閱讀成就之跨層次影響: 以 PIRLS 2006 為例。教育科學研究期刊, 56(2), 69-105.

臺灣 PISA 國家研究中心主編 (2018) .PISA2018 台灣學生表現. 心理出版社.

Abdulrahman, M. D., Faruk, N., Oloyede, A. A., Surajudeen-Bakinde, N. T., Olawoyin, L. A., Mejabi, O. V., ... & Azeez, A. L. (2020). Multimedia tools in the teaching and learning processes: A systematic review. *Heliyon*, 6 (11) .

<https://doi.org/10.1016/j.heliyon.2020.e05312>

Anne E. Mesmer, H., Hiebert, E. H., Cunningham, J. W., & Kapania, M. (2020). Does one size fit all? Exploring the contribution of text features, text content, and grade of use on comprehension. *Reading Psychology*, 42 (1), 42-72.

<https://doi.org/10.1080/02702711.2020.1861864>

B. A. Lively & S. L. Pressey, "A method for measuring the vocabulary burden of textbooks," *Educational administration and supervision*, vol. 9, pp. 389-398, (1923)

Collins-Thompson, K. (2014). Computational assessment of text readability: A survey of current and future research. *ITL-International Journal of Applied Linguistics*, 165 (2), 97-135.
<https://doi.org/10.1075/itl.165.2.01col>

Crossley, S., Heintz, A., Choi, J. S., Batchelor, J., Karimi, M., & Malatinszky, A. (2023). A large-scaled corpus for assessing text readability. *Behavior Research Methods*, 55 (2), 491-507.

Carretti, B., Toffalini, E., saponaro, C., Viola, F., & Cornoldi, C. (2020). Text Reading Speed in a Language with a Shallow Orthography Benefits Less from Comprehension as Reading Ability Matures. *British Journal of Educational Psychology*, 90, 91-104.

<https://doi.org/10.1111/bjep.12307>

Fuchs, K. (2023, May). Exploring the opportunities and challenges of NLP models in higher education: is Chat GPT a blessing or a curse?. In *Frontiers in Education* (Vol. 8, p. 1166682). *Frontiers Media SA*. <https://doi.org/10.3389/feduc.2023.1166682>

Gupta, A., Bansal, A., & Roy, K. (2021, May). Solar energy prediction using decision tree regressor. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 489-495. IEEE.

Huang, D., Yan, C., Li, Q., & Peng, X. (2024). From Large Language Models to Large Multimodal

- Models: A Literature Review. *Applied Sciences*, 14 (12), 5068. <https://doi.org/10.3390/app14125068>
- Joshi, N., Singh, G., Kumar, S., Jain, R., & Nagrath, P. (2020). Airline prices analysis and prediction using decision tree regressor. In *Data Science and Analytics: 5th International Conference on Recent Developments in Science, Engineering and Technology, REDSET 2019, Gurugram, India, November 15–16, 2019, Revised Selected Papers, Part I 5*, 70-186. Springer Singapore.
- Kincaid, J. P., Fishburne, L. R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, Fog Count and Flesch Reading Ease Formula) for navy enlisted personnel. Millington, TN: Navy Research Branch.
- Kulkarni, S., Katariya, J. K., & Potika, K. (2020, December). GloveNor: Glove for node representations with second order random walks. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 536-543. IEEE.
- Kalla, D., Smith, N., Samaah, F., & Kuraku, S. (2023). Study and analysis of chat GPT and its impact on different fields of study. *International journal of innovative science and research technology*, 8 (3).
- Levine, S. L., Cherrier, S., Holding, A. C., & Koestner, R. (2022). For the love of reading: Recreational reading reduces psychological distress in college students and autonomous motivation is the key. *Journal of American College Health*, 70 (1), 158–164. <https://doi.org/10.1080/07448481.2020.1728280>
- Murgia, E., Pera, M. S., Landoni, M., & Huibers, T. (2023, June). Children on ChatGPT readability in an educational context: myth or opportunity?. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization* (pp. 311-316).
- <https://doi.org/10.1145/3563359.3596996>
- OpenAI (2023). OpenAI Official Website. Introducing ChatGPT - Learn more. Available online at: <https://openai.com/blog/chatgpt/> (accessed February 10, 2023).
- Patel, N., Nagpal, P., Shah, T., Sharma, A., Malvi, S., & Lomas, D. (2023). Improving mathematics assessment readability: Do large language models help?. *Journal of Computer Assisted Learning*, 39 (3), 804-822. <https://doi.org/10.1111/jcal.12776>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532-1543.
- Smith, D. R. (1989). *The Lexile Scale in Theory and Practice*. Final Report.
- Sakketou, F., & Ampazis, N. (2020). A constrained optimization algorithm for learning GloVe embeddings with semantic lexicons. *Knowledge-Based Systems*, 195, 105628. <https://doi.org/10.1016/j.knosys.2020.105628>
- Schwarm, S. E., & Ostendorf, M. (2005, June). Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd annual meeting of the Association for Computational Linguistics (ACL'05)*, 523-530.
- Schleicher, A., Zimmer, K., Evans, J., & Clements, N. (2009). *PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science*. OECD Publishing (NJ).
- Thorndike, E. L. (1921). *The teacher's word book*. New York: Teachers College, Columbia University
- Weng, S. Y., Tseng, H. C., Sung, Y. T., & Chen, B.

- (2019, October). 基於階層式編碼架構之文本可讀性預測 (A Hierarchical Encoding Framework for Text Readability Prediction) . In Proceedings of the 31st Conference on Computational Linguistics and Speech Processing (ROCLING 2019) (pp. 334-342) .
- Vajjala, S., & Meurers, D. (2012, June). On improving the accuracy of readability classification using insights from second language acquisition. In Proceedings of the seventh workshop on building educational applications using NLP (pp. 163-173) .<https://doi.org/10.3115/v1/D14-1162>
- Wang, Q., Wang, P., & Chang, Y. (2023) . Deep Learning-Based Intelligent Image Recognition and Its Applications in Financial Technology Services. *Traitement du Signal*, 40 (2) .
<https://doi.org/10.18280/ts.400233>
- Wang, S., Zeng, Q., Ni, W., Cheng, C., & Wang, Y. (2023) . ODP-Transformer: Interpretation of pest classification results using image caption generation techniques. *Computers and Electronics in Agriculture*, 209, 107863.
<https://doi.org/10.1016/j.compag.2023.107863>
- (N.d.) . Project Gutenberg.
<https://www.gutenberg.org/>